Graphic Overlays in High-Precision Teleoperation: Current and Future Work at JPL

Daniel B. Diner, Ph.D.
Steven C. Venema, B.A.
Automated Systems Section
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

ABSTRACT

In high-precision teleoperation, high-resolution visual depth information may be critical, thus requiring vision system capabilities quite different from lower precision teleoperation vision systems. Several possible approaches to providing this depth information are available. Multiple-camera television systems, 3-D television systems, and 3-D video graphics systems all have advantages and disadvantages.

Multiple camera TV systems provide depth information by providing several views of the workspace. In such systems, camera mobility is desirable. However, moving cameras can confuse the operator. Therefore, the operator must know at all times the location of each camera. Providing such information can be cumbersome and increase operator workload.

Converged stereo TV cameras configured for high-depth precision can yield significant depth distortions, thus making many high-precision tasks extremely difficult, even for trained operators.

Video graphic systems can provide depth information through a variety of techniques including monocular depth labeling by color, brightness, perspective, occlusion, etc., as well as traditional 3-D binocular image presentation. However, video graphics systems have a problem which TV systems do not have; i.e., when viewing unpredictable situations, graphics systems may not be able to provide critical information in a timely manner.

In space teleoperation additional problems arise, including signal transmission time delays. These can greatly reduce operator performance.

Recent advances in graphics open new possibilities for addressing these and other problems.

At JPL, we are currently developing a multi-camera system with normal and 3-D TV and video graphics capabilities. Trained and untrained operators will be tested for high-precision performance using two force-reflecting hand controllers and a voice recognition system to control two robot arms and up to 5 movable stereo or non-stereo TV cameras. Through extensive experimentation, we plan to evaluate a number of new techniques of integrating TV and video graphics displays to improve operator training and performance in teleoperation and supervised automation.

INTRODUCTION

Video graphics has recently advanced quite rapidly. Today, high-resolution, real-time graphic systems can be purchased off the shelf, thus establishing graphics as a candidate for real-time video image enhancement in high-precision teleoperation.

As the fields of robotics and teleoperation continue to develop, an increasing number of tasks which currently must be performed manually will be performed either remotely or under automation. Video graphics, a display technique for human observers, will most probably extend the capabilities of teleoperation more than robotics.

Graphics will be very useful to remote operators by providing information which would otherwise not be readily available, such as camera locations, repair manual diagrams, visual depth information, velocities of relevant objects on a video monitor, force-torque diagrams, etc.

In space, many of the tasks which are currently performed by EVA (extra-vehicular activity) will be performed in the future by IVA (intra-vehicular activity). This makes teleoperation and robotics extremely interesting to NASA.

Also, current EVA tasks which have traditionally been labeled as future robotic tasks may be accomplished sooner in the future under graphics-aided teleoperation. As time passes, the "division of labor" between robotics and teleoperation will be more clearly defined.

In this paper, we describe the future vision system of the Man-Machine Systems Research Lab at JPL. This lab is not to be confused with the Telerobot Demonstrator Testbed, described elsewhere in this conference.

BACKGROUND

When viewing a work space remotely, through a TV camera, the one imprecisely-displayed dimension is depth (i.e., distance from the TV camera.) This dimension is a critical requirement for good teleoperation.

Much work has been done on presenting the video depth information in 3-D stereo (1 - 10). High-precision, close-up, 3-D TV has been shown to have a depth-resolution/depth-distortion/image-alignment trade-off (7).

An alternative to 3-D TV is the use of a multiple-camera viewing system, where the depth information can be figured out by the operator by looking at the work space from several views simultaneously.

A third alternative is to use graphics information to provide depth information (10). Combinations of the above three depth display techniques are also feasible. Multiple 3-D views with graphics overlays promise to be very useful in teleoperation.

DISCUSSION

Current Work at JPL

Over the past 3 1/2 years, we have studied 3-D TV, both mathematically and experimentally. We have quantified the depth distortions, both for still and moving stereo camera rigs (7). We have found an optimal method of moving the stereo camera rig to minimize 3-D depth distortions caused by camera motions (8).

We have also demonstrated a stereo image presentation technique which yields aligned images, high depth resolution and low depth distortion, thus solving the trade-off problem (9,11). NASA has a patent on this technique.

Future Work at JPL

Although our stereo image presentation technique promises to enhance high-precision 3-D TV, multiple-camera viewing systems may still have an important role to play in the future of teleoperation, particularly with the addition of graphics. Single-camera stereo systems (11) provide the possibility of multiple stereo 3-D views.

We are building an experimental telerobotic work station with two robot arms surrounded by 5 movable TV cameras. The cameras will each be mounted on a computerized gantry frame, with one camera on each of the five sides of the gantry frame. That is, the front, the back, the left, the right and above. Each camera will have the ability to move in its plane (up-down and front-back for the two sides, up-down and left-right for the front and back, and left-right and foward-back for the top camera). In addition, each camera will be able to pan, tilt and change the power of the lens (zoom). Thus each camera can view the work space from any location and angle in its range of motion in its plane. Camera motions may be commanded by the operator, for example, using voice control, or may be automated, following the robot grippers as they move about the work space. We envision automating the system to tailor camera motions to the current task at hand.

Up to five monitors will be available for the five camera views. Two additional monitors may be available for system information, trouble shooting, etc. An image enhancement system will also be present which will include graphics capabilities, and perhaps image processing capabilities. The operator will be able to command (by voice control) which camera view will be displayed on each monitor. Initial configuration may be fixed, for example left camera on the left monitor, etc. This however is not required.

Our approach is both theoretical and experimental. The critical question, as always in this work, is operator performance. Experimentation alone can answer if operators perform better under one set of conditions than another. We intend to address a variety of topics in our research, including the following.

1. Camera Locations and Apparent Motion

When viewing a workspace with movable cameras, an operator can be greatly confused by not knowing at all times the locations, orientations, and motions of each camera. Apparent motion, when one believes that the world is moving when actually the camera is moving, is particularly confusing. When multiple cameras are available, the additional problem arises of knowing which camera view is presented on the monitor (or each monitor if there are multiple monitors). Graphics can help solve these problems by providing the necessary information.

We envision presenting a camera's video image with overlayed graphics information showing the location and orientation of the TV camera on the monitor. See Figure 1.

In Figure 1, the TV camera image shows the right robot holding a ball and the left robot holding nothing. In addition to the TV camera view is a top-view graphics image of the camera frame, showing the positions and orientations of the camera. In this configuration (top view) it is necessary to specify the height of the camera, perhaps with 3-D stereo depth, or some other form of depth labeling. The pan of the camera is obvious, and the tilt can be displayed graphically by lines and circles. For example, lines can mean 15 degrees elevation (front of camera above back) and pairs of circles can mean 15 degrees downward elevation. In Figure 1, the camera is tilted 45 degrees upward.

This graphics presentation can also be displayed on a separate monitor.

The advantage of this presentation is that although both robot grippers seem to appear at equal height, the fact that the camera is tilted upward tells us that, in fact, the left robot gripper is actually higher than the right gripper. Because we know that the camera is tilted 45 degrees

upward, we can judge better what motion will be necessary to hand the ball from the right robot gripper to the left robot gripper. If the TV image is stereo, then one can also judge the length of the motion.

Circles and lines need not be the best graphic illustration of tilt and, in fact, one of the variables we plan to research is how to best present the camera locations. "Best" is measured with respect to operator performance under a variety of tasks.

Another variable is the point of view of the graphics camera frame. In Figure 1, if the frame were presented from the side view, instead of the top view, both of the camera's translational degrees of freedom would be specified without depth labeling. Although this seems to be an obvious improvement, it may not be so. The top view unambiguously specifies which camera we are viewing through. In addition, with multiple monitors, operators may prove to perform better if all camera locations are specified from the same view.

Another alternative is to present all the cameras' locations on one graphics display of the camera frame. This would allow operators to use the graphics information and the voice controller to move a camera before viewing through it, thus saving valuable operator time. When using a system with several cameras, but only one monitor, moving a camera before viewing through it can be particularly valuable.

In a system where the lighting is variable, that is lights can be moved or turned on and off, the graphics can be used to specify the current state of the lighting system. The lighting can then be adjusted by voice control. In fact, any variable part of the system can be so specified, and adjusted.

Eventually, we plan to automate the system to control the cameras and graphics to provide the optimal view for each task during operation.

2. Image Jitter During Camera and Robot Motion

One may find it desirable for the camera to track the end-effector of the robot during robot motion. This raises the question of image jitter. Quite simply, if the camera does not move smoothly enough, or if the camera is not synchronized with the robot motion, the image of the robot will jitter on the monitor. Jittering images not only make precision operation difficult (one may want to tighten a bolt as the robot moves a unit across the workspace), but can increase operator discomfort.

We have designed our robot gantry so that the cameras can track the robot without jitter, provided only panning and tilting camera motions are used in the tracking. The maximum speed for jitter-free tracking is about 15 degrees/second. In our work configuration, that translates to robot motions of 15 to 70 cm/sec, depending on which camera is being used for tracking and the zoom setting of the lens. Thus, our system promises to provide excellent robot tracking capabilities.

3. Camera Motions and Coordinate Transformations

In a teleoperator work station, where movable cameras are viewing the work space, any panning, rolling or tilting of the cameras causes a mis-alignment between the coordinate system of the camera and the coordinate system of the operator viewing the monitor. For example, if the camera rotates 15 degrees to the left, the "straight ahead" direction on the monitor will actually be 15 degrees to the left. If one pushes a robot hand controller "forward", the robot will move foward, but will be seen on the monitor to move at an angle of 15 degrees to the right. This requires the operator to mentally transform coordinates continually, during operation, thus causing an increase in workload as well as an increase in the probability of operator error. If several movable cameras are presenting their images to several monitors, each may require a different coordinate transformation. The resulting increase in workload and

probability of operator error may well become unmanageable and dangerous.

When viewing a workspace with a movable camera, at least 7 coordinate systems exist: the Real World, the Work Space, the Robot Base, the Robot Joint, the Camera, the Control Station, and the Operator coordinate systems.

The problem then is to minimize operator workload produced by the transformations between these coordinate systems.

If the Robot-Camera Table is mounted on a moving vehicle, such as a planetary rover, then the Real World and the Work Space coordinate systems are different. If, however, the robot-camera table is not movable, then the Real World and the Work Space coordinate systems are equal. If the robot can move its base on the robot-camera table, then the Work Space and the Robot Base coordinate systems are different. If, however, the robot cannot move with respect to the robot-camera table, then the Work Space and the Robot Base coordinate systems are equal.

We use the term "Robot Base" coordinate system to distinguish from the Robot Joint coordinate system which customarily means the joint angles of the robot, and is different from the spatial (X,Y,Z,Pan,Tilt,Roll) coordinate system as defined from a fixed point on the robot, such as the robot base. The Robot Joint coordinate system is transformed to and from the Robot Base coordinate system by the software that controls the Robot, and is used by the robot's internal controller to move the robot joints correctly. Therefore we need not concern ourselves with the Robot Joint coordinate system here.

The Camera coordinate system is defined by what the camera sees. Thus, a camera panned to face southeast sees southeast as straight ahead. A camera rolled 180 degrees sees the earth as "up" and the sky as "down."

The Control Station coordinate system is defined with respect to the operator control station. Thus if the camera faces 15 degrees to the left in the Work Space coordinate system, then the direction straight ahead in the Work Space would be presented at 15 degrees to the right in the Control Station coordinate system.

The Operator coordinate system is defined with respect to the "subjective straight ahead" direction of the operator. A great deal of study has been conducted on this phenomena (12). For simplicity, let us assume that our operator defines this direction with respect to the operator control station, that is, the operator aligns himself or herself to face the control station directly. For now, we shall ignore the possibility that an operator may sit at an angle to the control station and not realize it.

At this point, let us consider a non-movable robot-camera table with a robot whose base is fixed to the table. Then the Real World, Work Space, and Robot Base coordinate systems are equal.

Our concerns then become the transformations between the Robot Base, the Camera, the Control Station, and the Operator coordinate systems. Let us see how they interact.

When a camera moves, say 15 degrees pan to the left, the Robot Base, the Control Station, and the Operator coordinate systems do not change. Only the Camera coordinate system changes; that is, straight ahead on the camera is now 15 degrees to the left for the Robot Base, the Control Station, and the Operator. Thus, motions directly away from the camera (directly into the monitor) are 15 degrees to the left w.r.t. all the other coordinate systems.

We believe that we have found a solution to the coordinate transformation problem, using graphics. JPL and NASA are currently considering patent rights on this method, and thus we cannot discuss it. If our idea is truly a solution, then its application will give the Robot Base, the Camera, the Control Station, and the Operator coordinate systems the same orientations. No transformations will need to be made by the operator, and no camera angles will need to be

remembered.

4. Orthogonal and Perspective Camera Views

We shall test operator performance with both orthogonal multi-camera views and perspective camera views. Let us discuss first the orthogonal-camera configuration.

Consider 3 cameras, one looking from above, one from one side, and one from the front. Consider 3 monitors placed with the top view above the front view, and the side view alongside the front view. This is the TV approximation to the classic orthogonal projection of mechanical drawings.

We say "the TV approximation" because it will not give a true orthogonal projection. In a TV image, two lines overlap if they point directly toward the camera, but in orthogonal projections, two lines overlap if they are perpendicular to the projection. See Figure 2. Thus, in fact, only the central line of view in the side camera is truly orthogonal to the front camera view. For the rest of the image, equal depth must be inferred. Two objects at equal depth from the front camera will have their front edges overlap exactly in the side camera's view only if the front edges of the two objects are viewed exactly at the middle of the side camera. The images of all other pairs of objects at equal depth will not overlap exactly. This is an important difference, and may prove to be the source of many operator errors when using orthogonal TV cameras. This point must not be overlooked, because it illustrates that orthogonal TV viewing may be misleading, particularly to people accustomed to orthogonal mechanical drawings, because they expect overlap to mean equal depth.

Let us now consider perspective viewing. This is the depth-display technique of the great Renaissance artists.

The left-brain/right-brain dichotomy between people suggests that people fall into analytic and artistic categories, particularly in terms of perception and motor performance. It also suggests that all of us have both artistic and analytic information processors in our heads. In any case, it is safe to say that we all have varying degrees of skill in judging depth both from orthogonal and perspective displays.

Unfortunately, orthogonal TV viewing has the problem discussed above. Thus, in multi-camera viewing, we may better perform using our perspective processor to judge depth. Surely, this needs to be tested experimentally, and carefully. We must first search for perspective views, and then test them against optimal orthogonal views.

5. Other Planned Graphics Overlay Experiments

We plan to test operator performance when aided by a variety of graphics overlays, including predictive displays and force-torque diagrams.

Predictive displays of robot positions are particularly useful when dealing with significant signal transmission time delays. When signals must travel long distances, for example through space, time delays between the time of an event and the time one views the event become significant. In a feedback loop, such as long-distance teleoperation, the time delay can greatly reduce performance.

Consider a teleoperated servicer (with a robot arm) on the moon, which is being controlled from earth. A time delay of about 4 seconds round-trip from the earth to the moon and back can be expected. Suppose at time t = 0, an operator moves the hand controller. At time t = 2 seconds, the servicer receives the signal and initiates the motion. At t = 4 seconds, the servicer is first seen to move on the operator's monitor.

With a predictive display, the expected final position of the robot arm is displayed as a graphics overlay on the monitor immediately after the hand controller is moved. This has been

described in detail elsewhere (13 - 16). For large time delays, the predictive display has been shown to improve operator performance (13). For small time delays, the extra information on the monitor from the predictive display may clutter the image and reduce operator performance. We shall test this question for a variety of tasks.

Force-torque displays graphically show the forces and torques sensed by the robot (17), at, say, the wrist. We shall test operator performance while varying the locations, size, and other presentation characteristics of the display. For example, we plan to overlay each robot's force-torque display on its forearm surface seen in the TV monitor. We shall also present the display on another monitor. Our goal, as in all our work, is to present the relevant information to the operator in a manner which increases operator performance.

CONCLUSION

Recent advances in graphics now make graphics a useful tool for enhancing video displays in teleoperation. At JPL, we are currently building a multi-camera viewing system with graphics capabilities. We plan to address certain problems in teleoperation that, once resolved, promise to enhance the capabilities of teleoperation. Our goal is to maximize the utility of teleoperation in space applications.

ACKNOWLEDGEMENTS

REFERENCES

1. Shields NL, Kirkpatrick M, Malone TB and Huggins CT: *Design Parameters for a Stereoptic Television System based on Direct Depth Perception Cues.* Proceedings of the Human Factors Society 19th Annual Meeting (1975), Washington, D.C.

2. Grant C, Meirick R, Polhemus C, Spencer R, Swain D and Tewell R: *Conceptual Design Study for a Teleoperator Visual System Report.* Martin Marietta Corporation Report, NASA CR-124273 (1973), Denver, Colorado.

3. Upton HW and Strother DD: *Design and Flight Evaluation of a Helmet-mounted Display and Control System.* In Birt JA & Task HL (Eds.): "A symposium on visually coupled systems: Development and application", Brooks Air Force Base Report AMD-TR-73-1 (1973).

4. Zamarian DM: *Use of Stereopsis in Electronic Displays: Part II − Stereoscopic Threshold Performance as a Function of System Characteristics.* Douglas Aircraft Company Report MDC J7410 (1976).

5. Pepper, RL, Cole RE, and Spain, EH: *The Influence of Camera Separation and Head Movement on Perceptual Performance under Direct and TV-displayed Conditions.* Proceedings of the Society for Information Display (1983).

6. Bejczy, AK: *Performance Evaluation of Computer-aided Manipulator Control.* IEEE International Conference on Systems, Man and Cybernetics (1976).

7. Diner DB and von Sydow M: *Stereo Depth Distortions in Teleoperation.* JPL Publication 87-1 (1987), Jet Propulsion Laboratory - NASA, USA. National Technical Information Service # 87N18985; Proceedings of the Twenty-Second Annual Conference on Manual Control (1986); AFWAL-TR-86-3093, Wright-Patterson AFB Aeronautical Labs, Ohio, USA.

8. Diner DB and von Sydow M: *Dynamic Stereo Vision Depth Distortions in Teleoperation.* Proceedings of the International Symposium on Teleoperation and Control (1988), Springer - Verlag, New York. ISBN 0-387-50054-5.

9. Diner DB and von Sydow M: *Static Stereo Vision Depth Distortions in Teleoperation,* "Ergonomics of Hybrid Automated Systems I" (1988), Elsevier Science Publishers B. V., Amsterdam.

10. Fisher SS, McGreevy, Humphries H, Robinett W: *Virtual Environment Display System* NASA Ames Research Center, IEEE Computer Graphics 21:1 (1987).

11. Diner DB and Fender DH: *Handbook of Human Engineering in Stereoscopic Viewing Devices* (In preparation). JPL External Publication, Jet Propulsion Laboratory, Caltech, Pasadena, California, USA. Expected Publication 1989.

12. Howard IP: *Perception of Posture, Self Motion and Visual Vertical Handbook of Perception and Human Performance I.* John Wiley and Sons, Inc., New York (1986).

13. Hashimoto T, Sheridan TB and Noyes MV: *Effects of Predictive Information in Teleoperation with Time Delay.* Japan Journal of Ergonomics 22,2 (1986).

14. Noyes MV and Sheridan TB: *A Novel Predictor for Telemanipulation Through a Time Delay.* Proceedings of Twentieth Annual Conference on Manual Control (1984), NASA Ames Research Center, Moffett Field, California.

15. Buzan FT: *Control of a Teleoperator in the Presence of Large Time Delays.* MIT Man-Machine Systems Laboratory, JPL Contract no. 956892, Mid-Term Reports, June (1986) and May (1987).

16. Venema SC and Bejczy AK: *The Phantom Robot: Predictive Displays for Teleoperation with Time Delay.* IEEE International Conference on Robotics and Automation (1989). Submitted.

17. Bejczy AK and Dotson RS: *A Force-torque Sensing and Display System for Large Robot Arms.* Proceedings of IEEE Southeastcon82, Destin, Florida (1982).
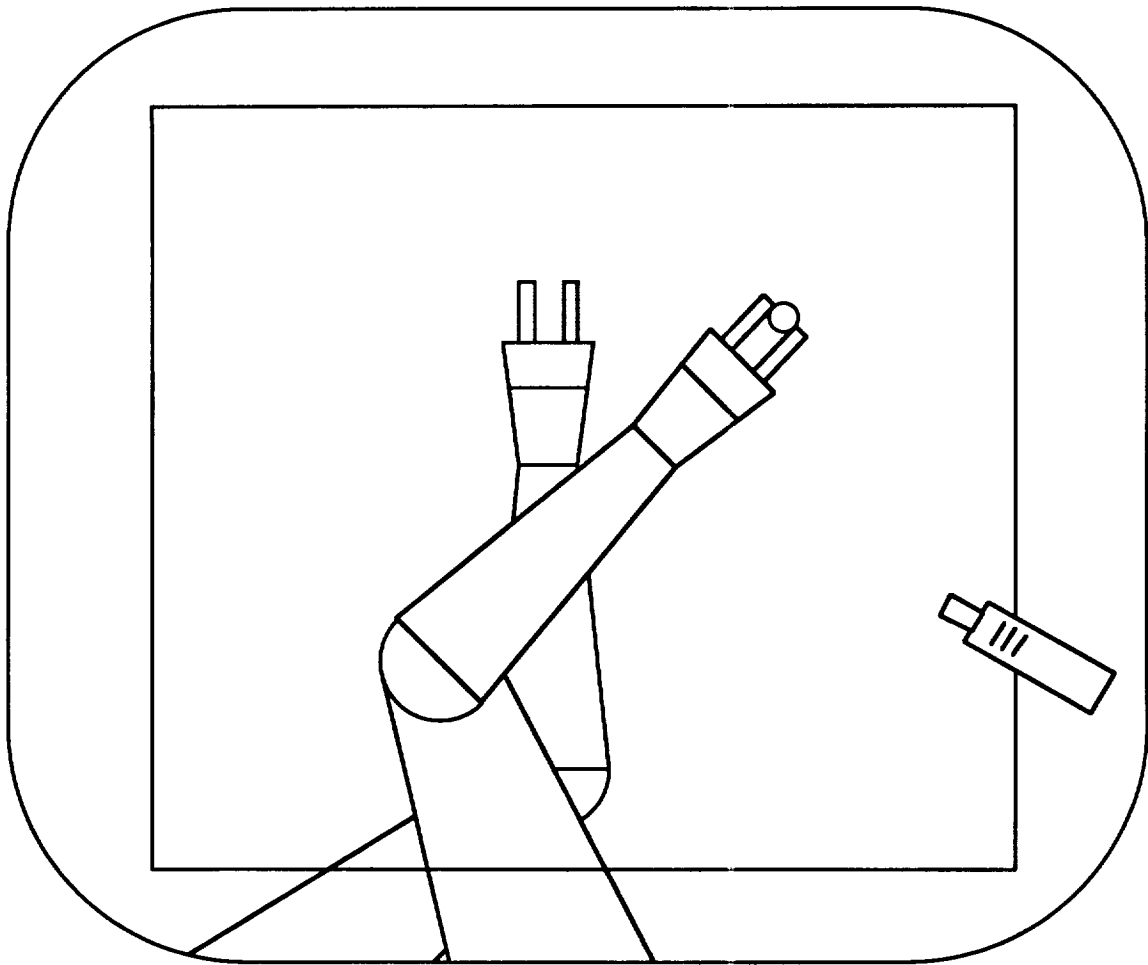
Figure 1: TV camera image of two robot arms with graphic overlay
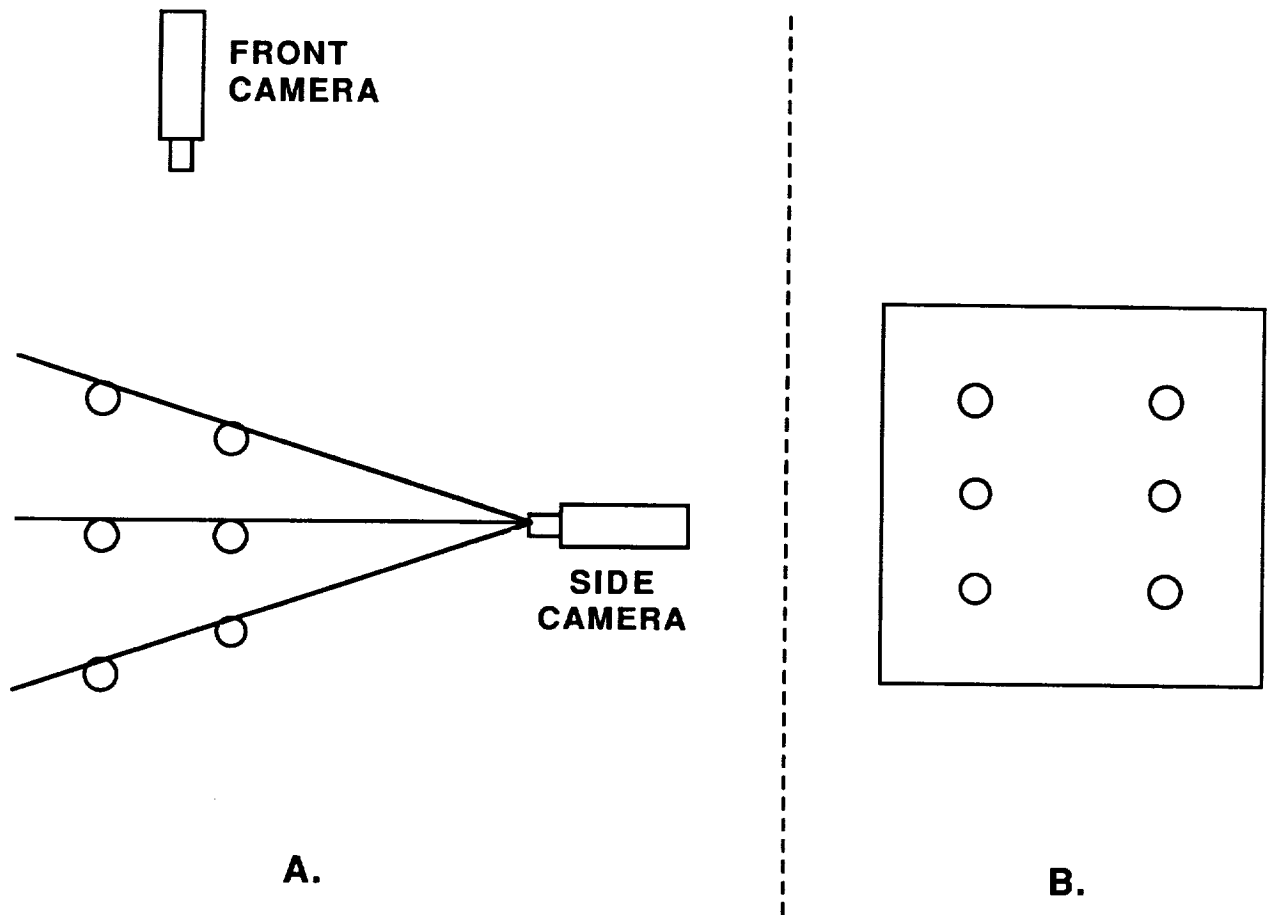of top-down view of camera frame and camera location.

Figure 2: Top view of the locations of 3 pairs of objects which:
    (a) overlap in a side TV-camera view, and
    (b) overlap in a side view in standard orthogonal
        mechanical drawings.